

Statistics

Summer Term 2021

Massimiliano Canzi

March 2021

1 Summary

1.1 Instructors

Max (*short for Massimiliano*) **Canzi** | <massimiliano.canzi@uni-konstanz.de>
Room G114 | Office hours over appointment

Chen-An Chang (TA) | <chen-an.chang@uni-konstanz.de>
Room H134 | Available for clarifications over e-mail

1.2 Requirements

Understanding of fundamental concepts in *descriptive* statistics (e.g. mean, median, mode, ...). Some understanding of basic programming concepts (e.g. variables, functions).

1.3 Schedule

- **Lecture:** asynchronous, posted on Ilias on Friday mornings at 10:00 am
- **Tutorials:** Tuesdays 9-10, Wednesdays 14:15-15:00 (*Starting Week 2*). Students are encouraged to attend the slot that suits them better, granted they attend at least one per week. Students are also encouraged to attend both sessions should they want to.

1.4 Textbook

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

1.5 Original outline

This course offers an introduction into state-of-the-art statistical methods for the quantitative analyses of linguistics data with practical examples (using the Software R), including linear regression, logistic regressions, linear-mixed effects modelling, etc.. By the end of the course, students will be able to decide on the best course of action for the quantitative analysis of their own linguistic data and they will have also learnt the best practices of reporting their own findings (e.g. data visualisation, test statistics) and correctly interpreting other people's.

1.6 Assessment

- **Homework** is assigned weekly. Homework is mostly carried out by students independently. Students will have the chance to ask questions during tutorials.
- **Midterm**: a coding exercise, similar to homework assigned during previous weeks, that students will have to hand in. Date to be determined. The midterm is worth 30% of total grade.
- **Paper**: Students will have to write a 2500-word paper on the analysis of a complex dataset (provided by the instructor). The paper will describe the methods used for the analysis as well as the main hypotheses and findings of the study. The analysis of the data set can be carried out in groups but each student will have to write their own paper.

2 Contents

The Statistics seminar combines theoretical lectures and hands-on coding tutorials to provide a fundamental understanding of both topics in statistics as well as skills necessary to be able to process, analyse and visualise linguistic data.

2.1 Data modelling

- Descriptive statistics (measures of central tendency and dispersion)
- Data types
- Introduction to regression modelling
- Linear mixed-effect models (LMEM)
- Generalised linear mixed-effect models (GLMEM)
- Fixed and random effects in linear modelling
- Interpreting main effects and interactions
- Shortcomings of linear modelling

2.2 R: Applying theoretical knowledge to real data

- Understanding the R and R Markdown philosophy
- Processing data in R
- Statistical modelling in R: best practices and examples
- Fundamentals of data visualisation
- Visualising raw data vs model estimates